*Proc. Natl. Acad. Sci. USA*
Vol. 88, pp. 11076–11080, December 1991
Biophysics

# A strategy for finding classes of minima on a hypersurface: Implications for approaches to the protein folding problem

(nonlinear optimization/"antlion")

TERESA HEAD-GORDON, FRANK H. STILLINGER, AND JULIO ARRECIS*

AT&T Bell Laboratories, Murray Hill, NJ 07974

*Contributed by Frank H. Stillinger, September 13, 1991*

**ABSTRACT** **Locating the native structure of a given protein is a task made difficult by the complexity of the potential energy hypersurface and by the huge number of local minima it contains. We have explored a strategy (the "antlion" method) for hypersurface modification that suppresses all minima but that of the native structure. Transferrable penalty functions with general applicability for modifying a hypersurface to retain the desired minimum are identified, and two blocked oligopeptides (alanine dipeptide and tetrapeptide) are used for specific numerical illustration of the dramatic simplification that ensues. In addition, an intermediary role for neural networks to manage some aspects of the antlion strategy applied to large polypeptides and proteins is introduced.**

## Section 1. Introduction

The protein folding problem is one of the most significant and intriguing challenges in molecular biophysics (1–3). The native forms that have been determined for naturally occurring proteins display a fascinating variety of three-dimensional structures exquisitely tailored to biological function. In addition, the experimentally observed folding kinetics of naturally occurring proteins involves time scales on the orders of microseconds to minutes, rather than the millennia expected for random-walk searches among conformational alternatives (4). General principles by which any linear sequence of amino acid residues encodes information about the native structure, and the most efficient kinetic pathway to this structure, still remain largely out of reach. The present paper is devoted to the exploration of a strategy that, we hope, will eventually illuminate these general principles.

From the theoretical viewpoint, the protein folding problem comprises three components. The first involves specifying the free energy (potential of mean force) hypersurface for arbitrary configurations of a given polypeptide immersed in the solvent of interest. The second concerns the kinetic pathway by which any non-native structure (in particular that of the newly synthesized protein emerging from the ribosome) manages to attain the native structure. The third amounts to nonlinear optimization on the free energy hypersurface to identify the native structure and any feasible alternative folding structures (low-lying relative free energy minima) and to show how they are determined by the amino acid sequence and solvent. In relation to this final point, we note that there is some question as to whether the native structure is always the global free energy minimum (5, 6) or a very long-lived metastable state (7). For the present we concentrate on the last of these three components, assuming that at least an approximation to the relevant free energy hypersurface is available.

Chemically realistic approximations to the conformational behavior of polypeptides inevitably entail hypersurfaces with enormous complexity. It is generally believed that $\Omega$, the number of distinct local minima, rises approximately exponentially with $N$, the number of residues:

$$\ln \Omega \cong \alpha N. \qquad [1.1]$$

A rough range of $N$ for naturally occurring proteins is 100 to 1000, while $\alpha$ probably lies in the range of 1 to 10. Searching for the native structure among such a large number of candidates is daunting to say the least.

Theoretical and computational strategies for solving the native protein minimization problem have been quite varied. Some examples include brute force minimization (8), statistical mechanical models ranging from that of Zimm and Bragg (9) to Monte Carlo simulations of highly simplified lattice models (10), and the application of neural network concepts to prediction of protein secondary structure (11–14) and tertiary structure (15–18). The adaptation of spin-glass theory to associative memory Hamiltonians for proteins (15–17) and explicit neural network training on distance matrices (18) offer promise for overcoming the deficiencies of traditional neural network implementations (11–14), where only a maximum of $\approx$67% reliability has been achieved for prediction of secondary structure.

This paper reports results of an exploratory investigation that was undertaken to determine the applicability of a general optimization strategy to the protein folding problem, the so-called "antlion" method (19, 20). This approach relies on the ability to deform the objective function hypersurface in such a way that the basin surrounding the global minimum (or a metastable minimum) widens and dominates. It takes its name from the family of subterranean insects that lie in wait at the bottom of victim-entrapping basins. In the present context it is the job of the antlion method to replace the complicated protein hypersurface by one for which $\alpha = 0$ in Eq. **1.1**. Any elementary minimization routine such as steepest descent on the modified hypersurface would then automatically converge to the single remaining minimum type, which by construction should be identical to the global (or even a preselected metastable) minimum of the starting problem, or at least a close approximation thereto. The final step in the antlion method is to optimize on the undeformed hypersurface, using the converged structure derived from the simplified potential energy surface as an initial guess. We note that this same strategy is possible for classes of minima, where the complexity of the hypersurface is reduced to $0 < \alpha << 1$. Stillinger (21) and Piela *et al.* (22) have proposed the use of a diffusion equation method for deforming hypersurfaces to retain only the global minimum. The antlion method

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Chemistry Department, Massachusetts Institute of Technology, Cambridge, MA 02139.

differs from the diffusion equation method in several respects; the most important difference is that the diffusion equation method has been explicitly demonstrated on one- and two-dimensional model systems only (22), whereas the antlion method presented here is applicable to an arbitrarily large number of dimensions of biological interest (23).

To create a generally useful antlion strategy, at least part of the hypersurface modification algorithm must contain mathematical operations that are transferable between different polypeptides, and in particular from oligomers to higher molecular weight polypeptides. The success of transferability may ultimately allow insight into the nature of the pathways by which a random coil accesses the correct tertiary-structure minimum, in addition to identifying whether a protein is kinetically or thermodynamically stabilized. In this spirit we have focused considerable attention on the blocked alanine dipeptide (Fig. 1) and alanine tetrapeptide (Fig. 2). As discussed in more detail below, we foresee ultimately employing a neural network automation procedure to manage some aspects of the antlion modification.

*Section 2* presents the details of the objective potential function for both alanine dipeptide and alanine tetrapeptide, as well as the specific algorithms for searching conformation space for minima. *Section 3* reports energies and structures for the alanine dipeptide minima, identifies some transferable modifications of the original potential energy, and shows how the resulting modified potential is drastically simplified to one surviving minimum. In the same section, we extend these considerations to the alanine tetrapeptide, again demonstrating the capacity for dramatic simplification to a single minimum hypersurface. *Section 4* summarizes our results and outlines our best projection for future development of the antlion approach.

## Section 2. Methods and Models

**Potential Function.** A reasonable approximation to the hypersurface of an arbitrary polypeptide is the following well-established empirical potential energy function:

$$V_0 = \sum_i^{\text{bonds}} k_{bi}(b_i - b_{i0})^2 + \sum_i^{\text{angles}} k_{\theta i}(\theta_i - \theta_{i0})^2 + \sum_i^{\text{improper}} k_{\tau i}(\tau_i - \tau_{i0})^2 + \sum_i^{\text{torsions}} k_{\omega i}[1 + \cos(n_i\omega_i + \delta_i)] + \sum_{i<j}^{M}\sum^{M} \{C\, q_iq_j/r_{ij} + \varepsilon_{ij}[(R_{ij}/r_{ij})^{12} - 2(R_{ij}/r_{ij})^6]\}. \quad [2.1]$$

The first four terms provide the connectivity potential; the bond length, bond angle, and improper torsion deformations are represented as harmonic potential functions with force constants $k_b$, $k_\theta$, and $k_\tau$ and equilibrium values $b_0$, $\theta_0$, and $\tau_0$, respectively. The torsional potential is represented as a Fourier cosine expansion, where $k_\omega$ is the force constant, $\delta$ is the phase, and $n$ is a multiplicity factor that allows for inclusion of higher harmonics. While the chirality of the $\alpha$-carbon center of all amino acids except glycine dictates the use of a general Fourier series, the cosine series will be adequate for the current study. The empirical parameters used and the specific torsions evaluated (only one dihedral term is evaluated for rotation around a given bond) for both alanine dipeptide and alanine tetrapeptide are presented in Table 1 (24).

The remaining terms in Eq. **2.1** are nonbonded interactions, which are modeled as pairwise coulomb electrostatic and Lennard–Jones interactions. The $i, j$ sums are restricted to pairs of atoms separated by three or more intervening bonds between the pair. The Lennard–Jones interaction

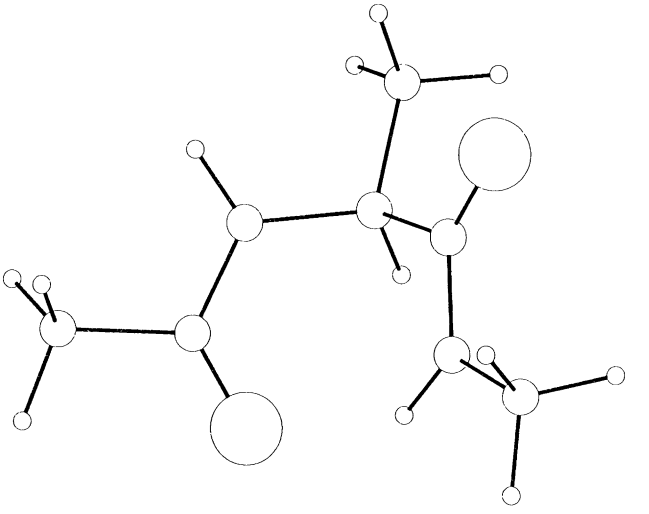

FIG. 1. Structure of the global-minimum conformer for the blocked alanine dipeptide.

parameters are evaluated using simple mixing rules of the individual atomic parameters:

$$\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$$

$$R_{ij} = (R_{ii} + R_{jj})/2. \quad [2.2]$$

In addition, the electrostatic interactions are scaled by a factor $C = 0.5$ when the pair under consideration are separated by exactly three bonds; otherwise $C = 1.0$. In Table 2 we list the Lennard–Jones parameters (24) used for both alanine dipeptide and alanine tetrapeptide. In Table 3 we provide the charges for alanine dipeptide (24), which differ slightly from the charges for alanine tetrapeptide in the same table, in order to ensure that charge neutrality is maintained. The set of the connectivity and nonbonded parameters in Tables 1–3 for the di- and tetrapeptide will henceforth be referred to as yielding the unmodified interaction, $V_0$. A description of the corresponding sets for the modified interactions is left to *Sections 3* and *4*.

**Characterization of Minima.** Given an objective function such as that in Eq. **2.1**, we require a method for obtaining a majority, or if possible all, of the minima on the hypersurface it represents. For simplicity we employ a Monte Carlo

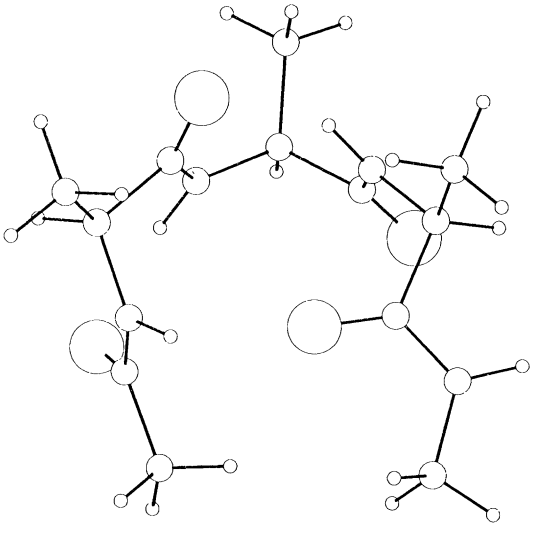

FIG. 2. Structure of the global-minimum conformer for the blocked alanine tetrapeptide.

Table 1. Parameters of the intramolecular potential energy function for the alanine dipeptide and tetrapeptide

| | Force constant | Equilibrium value |
|---|---|---|
| Bond type | $k_b$, kcal/(mol·Å$^2$) | $b_0$, Å |
| CT-HA | 340.0 | 1.090 |
| CT-C | 279.0 | 1.515 |
| C-O | 640.0 | 1.225 |
| C-N | 350.0 | 1.335 |
| N-H | 465.0 | 1.000 |
| N-CT | 310.0 | 1.460 |
| CT-CT | 268.0 | 1.515 |
| Angle type | $k_\theta$, kcal/(mol·rad$^2$) | $\theta_0$, deg |
| CT-C-O | 66.0 | 122.6 |
| CT-C-N | 64.0 | 113.9 |
| HA-CT-HA | 37.0 | 109.8 |
| HA-CT-C | 45.0 | 109.5 |
| C-N-H | 32.0 | 120.9 |
| C-N-CT | 44.0 | 117.6 |
| O-C-N | 98.0 | 125.0 |
| N-CT-HA | 46.5 | 109.2 |
| N-CT-CT | 76.0 | 111.0 |
| N-CT-C | 58.0 | 112.8 |
| H-N-CT | 23.0 | 120.8 |
| CT-CT-HA | 37.5 | 109.5 |
| Improper type | $k_\tau$, kcal/(mol·rad$^2$) | $\tau_0$, deg |
| C-CT-N-O | 125.0 | 0.0 |
| N-C-CT-H | 28.0 | 0.0 |
| Dihedral type | $k_\phi$, kcal/mol | $\delta$, deg ($n$) |
| CT-C-N-CT | 9.5 | 180 (2) |
| HA-CT-C-N | 2.2 | 0 (3) |
| HA-CT-N-C | 0.3 | 0 (3) |
| N-CT-C-N | 0.7 | 180 (2) |
| N-CT-CT-HA | 1.6 | 0 (3) |

CT corresponds to $C_\alpha$, $C_\beta$, CTR, and CTL. HA corresponds to $H_\alpha$, $H_\beta$, HTR, and HTL.

heating and quenching protocol, and subsequent minimization, to search exhaustively for minima of $V_0$ and of all of the modified functions discussed below. In some cases we use a minimization procedure with starting structures expected to be near stable stationary points.

The heating phase of any given Monte Carlo run consists of specifying an initial configuration of the atoms of alanine dipeptide or alanine tetrapeptide and generating configurations at a temperature of 20,000 K by using the Metropolis algorithm (25). A step size in Cartesian space of ±0.125 Å for *every* atom at *every* step results in a 50% acceptance rate for the 500,000-step run. Configurations are sampled every 10,000 steps, resulting in 50 configurations, which are used as starting structures for the quenching portion of the Monte Carlo search.

Table 2. Parameters of the Lennard–Jones function for the alanine dipeptide and tetrapeptide

| Atom type | $\varepsilon_{ii}$, kcal/mol | $R_{ii}$, Å |
|---|---|---|
| HTL/HTR | 0.0450 | 1.468 |
| CTL/CTR | 0.0903 | 1.800 |
| C | 0.1410 | 1.870 |
| O | 0.2000 | 1.560 |
| N | 0.0900 | 1.830 |
| H | 0.0498 | 0.800 |
| $H_\alpha$ | 0.0450 | 1.468 |
| $C_\alpha$ | 0.0903 | 1.800 |
| $C_\beta$ | 0.0903 | 1.800 |
| $H_\beta$ | 0.0450 | 1.468 |

Table 3. Parameters of the electrostatic function for the alanine dipeptide and tetrapeptide

| | $q_{ii}$, $e$ | |
|---|---|---|
| Atom type | Dipeptide | Tetrapeptide |
| HTL/HTR | 0.0000 | 0.0000 |
| CTL/CTR | 0.0000 | 0.0000 |
| C | 0.5500 | 0.5500 |
| O | −0.5500 | −0.5500 |
| N | −0.3500 | −0.3500 |
| H | 0.2500 | 0.2500 |
| $H_\alpha$ | 0.1000 | 0.1000 |
| $C_\alpha$ | 0.0000 | 0.0000 |
| $C_\beta$ | −0.2600 | −0.2917 |
| $H_\beta$ | 0.1200 | 0.1083 |

The Monte Carlo quenching stage consists of generating configurations at 10 K using a step size of 0.0005 Å, again resulting in an acceptance rate of 50% for Metropolis sampling. The Monte Carlo quench is terminated after 30,000 steps, and a BFGS minimization algorithm (26) is then used to determine the closest stationary point.

## Section 3. Hypersurface Modification for Alanine Oligopeptides

**Alanine Dipeptide.** We provide an enumeration in Table 4 of all L minima found by the Monte Carlo/minimization protocol outlined in *Section 2*; we note that we have also found most of the mirror images (D form) of the entries in Table 4. A large majority of the minima correspond to structures where a cis–trans isomerization of one or both peptide groups has occurred. The remaining four minima not represented in the preceding class are those which are found by a search through the two-dimensional space defined by the internal coordinate torsions $\phi$ (C-N-$C_\alpha$-C) and $\psi$ (N-$C_\alpha$-C-N). Fig. 3 displays an energy contour map showing these four energy minima, which is generated by constraining $\phi$ and $\psi$ and allowing relaxation of the remaining degrees of freedom (27). The lowest-energy structure of the $\Omega \geq 35$ minima corresponds to the $C7_{eq}$ all-trans L conformer (Fig. 1), which may be described as a seven-membered ring closed by an intramolecular hydrogen bond, with the side-chain methyl group equatorial to the plane of the ring.

Table 4. Enumeration of the alanine dipeptide minima

| $\phi$, deg | $\psi$, deg | $\omega_1$ | $\omega_2$ | $E$, kcal/mol |
|---|---|---|---|---|
| −78.1 | 72.5 | trans | trans | −32.391 |
| 163.7 | −163.9 | trans | trans | −30.997 |
| −163.3 | 149.9 | trans | cis | −30.766 |
| 71.6 | −65.9 | trans | trans | −30.339 |
| 74.8 | −143.2 | trans | cis | −29.630 |
| −157.9 | 78.3 | trans | cis | −28.430 |
| −169.7 | −55.3 | trans | trans | −26.258 |
| −53.8 | −50.5 | trans | cis | −26.215 |
| 70.1 | 24.2 | cis | trans | −26.037 |
| −72.8 | 145.2 | cis | cis | −25.995 |
| 65.4 | 36.6 | trans | cis | −25.870 |
| 152.3 | −148.6 | cis | cis | −25.716 |
| −166.7 | −42.3 | trans | cis | −25.284 |
| 152.5 | −160.7 | cis | trans | −24.841 |
| 71.6 | −156.7 | cis | trans | −24.833 |
| −68.1 | −39.8 | cis | cis | −24.170 |
| −49.5 | −48.8 | cis | trans | −23.615 |
| 76.0 | 167.3 | trans | cis | −23.375 |
| 53.2 | −131.6 | cis | trans | −20.799 |
| 68.4 | −178.8 | cis | trans | −20.744 |

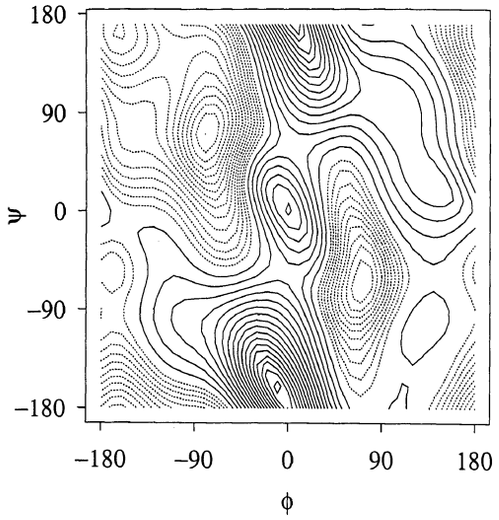


FIG. 3. The $\phi$, $\psi$ surface derived from the unmodified Hamiltonian, $V_0$ (Eq. **2.1**) for alanine dipeptide. The $\phi$ and $\psi$ variables are held fixed at each grid point (10° spacing), and all other degrees of freedom are relaxed. The dashed lines denote contours of 0.5 kcal/mol and extend from the zero of energy (the $C7_{eq}$ conformer) to 7.0 kcal/mol. Solid contours are drawn every 1.0 kcal/mol thereafter.

The first type of modification of $V_0$ is to eliminate all minima where one or both peptide groups are in the cis conformation (where exceptions are to be made if the residue is a proline) (1). We note that the peptide torsion potential used in Eq. **2.1** is specifically

$$V = 9.5[1 + \cos(2\omega + \pi)], \qquad [3.1]$$

which favors minima at both $\omega = 0$ and $\omega = \pi$. The obvious modification of Eq. **3.1** to favor the trans form is to change the multiplicity factor of 2 to 1, and to change the phase from $\pi$ to 0. To maintain the correct curvature at the minimum, we use a force constant of 38 kcal/mol, so that

$$V' = 38.0[1 + \cos(\omega)]. \qquad [3.2]$$

In addition, we will always desire the L configuration of a polypeptide sequence. To maintain the desired chirality, we incorporate an improper dihedral funotion,

$$V'' = 125.0(\tau - \tau_0)^2, \qquad [3.3]$$

for the torsions $C_\alpha$-N-C-$C_\beta$ ($\tau_0 = 33.0°$) and $C_\alpha$-N-C-$H_\alpha$ ($\tau_0 = -33°$). While the $V'$ and $V''$ modifications are trivial and in some sense physically unimportant in relevant areas of configuration space for biological molecules of interest, this serves as an illustrative example for what is to follow. For the case of alanine dipeptide, this modification permits us to visualize transforming the energy surface in Fig. 3 to retain only the $C7_{eq}$ conformer.

We have considered a number of modifications to the potential energy function representing the surface in Fig. 3 in order to retain only the global energy minimum. Our criteria for a successful modification are (*i*) that the penalty function explicitly or implicitly incorporate information about the tertiary structure of any peptide, (*ii*) that the functional form of the modification is transferable across any polypeptide sequence, and (*iii*) that a variety of conformations can be distinguished in a given segment of polypeptide ranging from the random coil to secondary structure conformers such as the $\alpha$-helix and $\beta$-sheet.

The generic penalty function

$$V''' = k_\phi[1 - \cos(\phi - \phi_0)] + k_\psi[1 - \cos(\psi - \psi_0)] \qquad [3.4]$$

fulfills the above objectives. For the case of polypeptides with minimal side chains such as glycine or alanine, the $\phi$, $\psi$ variables most directly define the tertiary structure; for polypeptides with more complex side chains, the same type of penalty function can be applied to the $\chi_i$ dihedrals as well, so that the functional form is transferable to any sequence of amino acid residues. Finally, for appropriately defined $\phi_0$ and $\psi_0$ parameters, it allows discrimination among the pool of relevant conformers observed in large polypeptides and proteins. This function (using the parameters defined in Table 5) indeed accomplishes the simplification of the surface in Fig. 3 to a single minimum: the global, $C7_{eq}$ minimum as exhibited in Fig. 4.

**Transferability to Alanine Tetrapeptide.** The number of unique minima for the alanine tetrapeptide case is quite large, even when cis peptides and D isomers are eliminated from consideration. However, the alanine tetrapeptide system offers some simplification for classifying these minima when one considers the conformational space of the tetrapeptide to comprise three sets of $\phi$, $\psi$ dihedrals. The alanine tetrapeptide shows the following minima in any given $\phi_i$, $\psi_i$ space: $C7_{eq}$, $C7_{ax}$, C5, $\alpha'$, $\alpha_R$, $\alpha_L$, and polyglycine II, and several "unusual" minima that occur infrequently relative to the preceding seven. Thus a large majority of minima fall into a classification where the three sets of $\phi$, $\psi$ variables can adopt any combination of the seven conformers $C7_{eq}$ (−75°, 75°), $C7_{ax}$ (75°, −75°), C5 (−165°, 165°), $\alpha'$ (−165°, −55°), $\alpha_R$ (−60°, −45°), $\alpha_L$ (60°, 60°), and polyglycine II (−80°, 150°). Enumeration of these possibilities indicates that the number of unique minima is approximately

$$\Omega \cong 7^3 + M, \qquad [3.5]$$

where $M$ is the small number of minima that do not fall into the above seven conformer classification ($M \cong 25$ in our search). In the case of stable minima for all possible combinations of the seven conformers (343), $\Omega \geq 368$ or $\alpha \cong 3$ (again, ignoring the possibility of cis and D conformers). We note that our search found most (but not all) of the $7^3 = 343$ simple possibilities, which can be attributed both to a lack of stability of a particular combination ($\alpha_R$, $\alpha_L$, $\alpha_R$, for example) and to the likelihood of incomplete sampling.

With these Monte Carlo results in hand for the unmodified tetrapeptide hypersurface, we then tested the transferability of the modification functions in Eqs. **3.2–3.4**. As before, the intention was to produce a modified potential surface possessing only a single minimum that corresponds closely to a preselected minimum of the complicated starting hypersurface. We have successfully achieved this goal in a manner that demonstrates considerable latitude in the character of the single minimum that is permitted to survive modification. Specifically, successful use of the transferable functions (with appropriate $\phi_0$, $\psi_0$ choices) has been demonstrated in the following independent cases: (*i*) retention of the global minimum, [$C7_{eq}$, $C7_{ax}$, polyglycine II], (*ii*) retention of a preselected metastable minimum, [$\alpha_R$, $\alpha_R$, $\alpha_R$], and (*iii*) retention of any one of the class of minima [$C7_{eq}$, $C7_{ax}$, *] where * denotes a "wild card" specification for the third $\phi$, $\psi$ pair (we have found this third $\phi$, $\psi$ pair on the unmodified surface to be either polyglycine II, $\alpha_R$, $C7_{eq}$, or $C7_{ax}$).

Table 5. Parameters for the alanine dipeptide potential, $V'''$

| Dihedral type | $k_\phi$, $k_\psi$, kcal/mol | $\phi_0$, $\psi_0$, deg | $n$ |
|---|---|---|---|
| C-N-$C_\alpha$-C | 7.5 | −75.0 | 1 |
| N-$C_\alpha$-C-N | 7.5 | 75.0 | 1 |

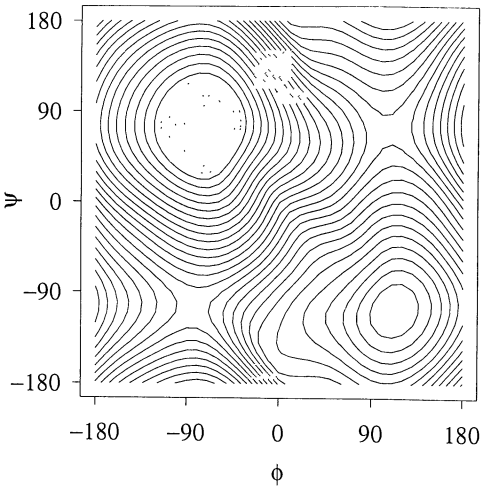


FIG. 4. The $\phi$, $\psi$ surface derived from the modified Hamiltonian (Eq. **3.1–3.4**) for alanine dipeptide. The $\phi$ and $\psi$ variables are held fixed at each grid point (10° spacing), and all other degrees of freedom are relaxed. The dashed lines denote contours of 2.0 kcal/mol and extend from the zero of energy (the $C7_{eq}$ conformer) to 8.0 kcal/mol. Solid contours are drawn every 3.0 kcal/mol thereafter.

## Section 4. Discussion and Conclusions

We have implemented a strategy for greatly simplifying peptide energy hypersurfaces in order to retain only one conformationally distinct minimum. To the extent that the surviving minimum corresponds closely to the desired minimum, the original conformational optimization problem undergoes drastic simplification. This approach has been illustrated by specific calculation for two small peptides, the blocked alanine dipeptide and tetrapeptide. For the former, 20 local minima (40, with mirror-image configurations) collapse to a single minimum upon application of suitable potential energy penalty functions. The functional forms of these penalty functions are immediately transferable to the tetrapeptide case (or indeed to larger polypeptides) and succeed in suppressing a much larger number of local minima on the starting hypersurface to favor, once again, a single surviving minimum. For both the dipeptide and the tetrapeptide, we have been able to arrange for the surviving minimum to closely approximate the desired minimum of the original surface, thereby simplifying the corresponding conformational search problem. These examples illustrate the promise for the extension of the antlion method to larger polypeptides and proteins, where determining the native structure from among the vast number of minima on the unmodified hypersurface is an intractable proposition, and therefore represents the case where simplification is highly desirable.

In the general context of protein conformational prediction, implementation of the antlion approach might at first glance seem to require at the outset knowledge of the secondary and tertiary structure sought. In particular, it would seem that sets of angles $\phi_0$, $\psi_0$ have to be identified to construct the necessary penalty functions; the $\phi_0$ and $\psi_0$ values used for short peptides would not necessarily transfer to longer peptides (28, 29). We foresee a fundamental role for neural networks (trained on a suitable protein database) to manage this aspect of the multidimensional optimization problem. However we hasten to stress that an important distinction exists between this intended role and that conventionally required of neural networks in the protein folding area. For the latter, the outputs of the network are the direct structure predictions, whether they concern secondary structure predictions (11–14) or residue contact-distance classification (18). The antlion method, however, would require only network predictions for the $\phi_0$, $\psi_0$ penalty parameters (and perhaps the corresponding force constants); subsequent minimization first on the modified potential hypersurface and then on the unmodified hypersurface serves as the tertiary predictor. Local violations of the neural network angle predictions become feasible, even likely, as the entire system seeks and finds its optimal final structure. In this respect our approach accommodates the presence of locally frustrated interactions in the interests of attaining a global minimum tertiary structure. It is the frequent occurrence of such intrinsic frustrated interactions that, in our view, has thus far limited the success rate of neural networks as direct predictors of secondary structure in proteins.

In summary, the simple-peptide-system results obtained thus far provide strong encouragement that the antlion strategy can be adapted to larger peptides and proteins. Indeed, we have been able to apply the antlion/neural network strategy, outlined in this section, successfully to the naturally occurring 26-residue polypeptide mellitin; details will be presented elsewhere (23).

J.A. is a recipient of an AT&T Cooperative Research Fellowship.

1. Gierasch, L. M. & King, J. (1990) *Protein Folding: Deciphering the Second Half of the Genetic Code* (American Association for the Advancement of Science, Washington).
2. Creighton, T. E. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5082–5086.
3. King, J. (1989) *Chem. Eng. News* **67,** 32–54.
4. Levinthal, C. (1968) *J. Chim. Phys.* **65,** 44–45.
5. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
6. Anfinsen, C. B. & Scheraga, H. A. (1975) *Adv. Prot. Chem.* **29,** 205–300.
7. Honeycutt, J. D. & Thirumalai, D. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3526–3529.
8. Scheraga, H. A. & Paine, G. H. (1986) *Ann. N.Y. Acad. Sci.* **482,** 60–68.
9. Zimm, B. H. & Bragg, J. K. (1959) *J. Chem. Phys.* **31,** 526–535.
10. Kolinski, A., Skolnick, J. & Yaris, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7267–7271.
11. Holley, L. H. & Karplus, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 152–156.
12. Bohr, H., Bohr, J., Brunak, S., Cotterill, R. M. J., Lautrup, B., Norskov, L., Olsen, O. H. & Petersen, S. B. (1988) *FEBS Lett.* **241,** 223–228.
13. Qian, N. & Sejnowski, T. J. (1988) *J. Mol. Biol.* **202,** 865–884.
14. Kneller, D. G., Cohen, F. E. & Langridge, R. (1990) *J. Mol. Biol.* **214,** 171–182.
15. Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93,** 6902–6915.
16. Wolynes, P. G. & Stein, D. L., eds. (1980) *Spin Glasses and Biology* (World Scientific, London).
17. Friedrichs, M. S. & Wolynes, P. G. (1989) *Science* **246,** 371–373.
18. Bohr, H., Bohr, J., Brunak, S., Cotterill, R. M. J., Lautrup, B., Norskov, L., Olsen, O. H. & Petersen, S. B. (1990) *FEBS Lett.* **261,** 43–46.
19. Stillinger, F. H. & Weber, T. A. (1988) *J. Stat. Phys.* **52,** 1429–1445.
20. Stillinger, F. H. & Stillinger, D. K. (1990) *J. Chem. Phys.* **93,** 6106–6107.
21. Stillinger, F. H. (1985) *Phys. Rev. B.* **32,** 3134–3141.
22. Piela, L., Kostrowicki, J. & Scheraga, H. A. (1989) *J. Phys. Chem.* **93,** 3339–3346.
23. Head-Gordon, T. & Stillinger, F. H. (1991) *J. Mol. Biol.*, in press.
24. Momany, F. A., Klimkowski, V. J. & Schäfer, L. (1990) *J. Comp. Chem.* **11,** 654–662.
25. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21,** 1087–1092.
26. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986) *Numerical Recipes* (Cambridge Univ. Press, Cambridge).
27. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963) *J. Mol. Biol.* **7,** 95–99.
28. Kabsch, W. & Sander, C. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1075–1078.
29. Argos, P. (1987) *J. Mol. Biol.* **197,** 331–348.